Research Policy 44 (2015) 849–861

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

How does public agricultural research impact society? A characterization of various patterns

A. Gaunand[a,b], A. Hocdé[c], S. Lemarié[a,b], M. Matt[a,b,*], E.de Turckheim[d]

[a] INRA, UMR1215 GAEL, F-38000 Grenoble, France
[b] University of Grenoble Alpes, UMR 1215 GAEL, F-38000 Grenoble, France
[c] Institut Francilien Recherche Innovation Société (IFRIS, unité INRA-SenS, Champs sur Marne, 77454 Marne-la-Vallée Cedex 2, France
[d] Delegation to evaluation, INRA Department head, 147 rue de l'Université, 75338 Paris Cedex 07, France

ARTICLE INFO

Article history:
Received 19 March 2014
Received in revised form 15 January 2015
Accepted 16 January 2015
Available online 24 February 2015

Keywords:
Public agricultural research
Societal impact
Impact pattern
Evaluation

ABSTRACT

This paper characterizes the various impact patterns generated by an agricultural public research organization (PRO), namely INRA (National Institute for Agronomic Research). We define an impact pattern as the combination of specific research outputs with specific actors that generates various types of impact. The analysis is based on information related to more than a thousand INRA innovations for which research outputs, beneficiaries, and impacts, have been codified. A classification based on the Partitioning Around Medoids (PAM) method is used to identify the seven main impact patterns.

There are two patterns that correspond to traditional INRA interventions to foster agricultural sector competitiveness; two that are related to innovations in health and economic issues; and two that have impacts on the conservation of natural resources. The seventh involves scientific advice related to public policy decisions. The research outputs and beneficiaries differ across these impact patterns. For example, those with economic impacts are more related to the agricultural sectors while impact patterns in the area of health affect industrial firms.

1. Introduction

Many countries aim at evaluating the societal impact of public research. This urge is explained by the shortages in public funds, the diffusion of New Public Management rules and by changes in the research system[1]. The public research system is increasingly governed by collaboration with industrial partners, interdisciplinarity, "context-driven" research targeted towards specific users and needs. Public Research Organizations (PROs) and researchers who formerly were evaluated only by their peers based on scientific excellence criteria, are being pushed by the different funding agencies and stakeholders to take a wider view of their performance that includes the societal utility of the knowledge produced. The PRO mission to achieve scientific progress has been extended to include the resolution of societal challenges through collaboration with and diffusion of knowledge towards socio-economic partners, and contribution to public policy decisions and scientific debate. In that context, PROs are required to adopt or develop methods that provide evidence of the societal returns from their research results. This pressure is especially strong for PROs involved in targeted research, e.g., in agriculture, to address stakeholders' issues. Evaluations of PROs have to show that the PROs' research results are generating various types of benefits for their various stakeholders and beneficiaries.

Such performance evaluations are made more complicated by the fact that: (i) the impacts are generally diversified, because PROs have multiple missions, (ii) the impacts generated result from the activities of multiple other actors than the PRO being evaluated, (iii) evaluation method needs to be standardized and applicable to various scientific domains and types of impact. Assessment of the broader impacts that take account of social, cultural, political, environmental, health, and economic returns, has been tackled by several studies and various evaluation exercises such as: the Research Excellence Framework in the UK (Martin, 2011); the NSF's broader impact criterion (Kamenetzky, 2013); the SIAMPI approach (Spaapen and Van Drooge, 2011); and the Payback Framework (Donovan and Hanney, 2011). These studies analyze the roles of multiple actors in the knowledge value and outcome generating processes. Many are based on case studies and combine qualitative data and quantitative metrics to assess the societal impacts of

* Corresponding author at: INRA, UMR1215 GAEL, F-38000 Grenoble, France. Tel.: +33 456525689.
E-mail address: Mireille.Matt@grenoble.inra.fr (M. Matt).

[1] cf. 'Mode 2' knowledge production (Gibbons et al., 1994), 'Entrepreneurial University' model (Etzkowitz, 1998).

http://dx.doi.org/10.1016/j.respol.2015.01.009
0048-7333/

research. These analyses are interesting to understand the mechanisms that explain the impact generating processes, but are limited in their ability to aggregate individual stories to understand the different impact patterns generated by a PRO. Our paper tries to fill this gap by using an original clustering method on a large dataset of innovations.

The objective of our paper is to characterize the diversity in the impact patterns generated by the French PRO, INRA (National Institute for Agronomic Research). We define an impact pattern as the combination of specific research outputs with specific actors that generates various types of impact. INRA conducts targeted research in agriculture, food, and the environment. Its research generates economic, health, environmental, and political impacts, all of which are taken into account in our impact patterns. INRA was created just after WWII, and has a long tradition of partnerships with socio-economic actors. Due to the development of private research in agriculture and the emergence of new societal demands, INRA has been repositioning its objectives towards more basic research, and some of the areas that prioritize public value (Bozeman and Sarewitz, 2011). Agricultural innovations based on research results generated by INRA usually involve multiple actors (extension services, private companies, government bodies).

The analysis presented in this paper relies on a dataset that includes over a thousand events and records of research results expected to generate societal impacts. Records were collected for every year from 1996 to 2010 through a bottom–up process. For each record, we codified the elements of the impact pattern: the outputs, the beneficiaries, and the potential impacts. A statistical analysis based on the PAM (Partitioning Around Medoids) clustering method allowed us to characterize seven impact patterns. These patterns reflect regularities in the diverse effects on society. An understanding of these patterns should allow INRA to improve its knowledge transfer practices. The value of our analysis is that it is based on a large sample of research events that is representative of the multiple activities and missions of the PRO. To the best of our knowledge, very few analyses have been conducted on such a large sample of research events.

The paper is organized as follows. Section 2 reviews the literature on the evaluation of public research impacts. Section 3 introduces the data that is used for the PAM analysis in Section 4. Section 5 presents the results and describes the seven impact patterns generated. Section 6 discusses the results and concludes.

2. Evaluating public research impacts

We first review the literature on the economic impacts, and then work on the assessment of their broader impact. This body of work helps us to identify the impact pattern components (actors, research outputs) that need to be considered. Finally, we present our research approach.

2.1. Evaluating the economic impacts

Numerous economic analyses are based on econometric estimations of an aggregate production function, and evaluate the contribution of research and development (R&D) to productivity growth at the country or industry level. In general, they find a positive relationship between public and private R&D expenditure and economic growth and productivity. These methods include approaches that calculate rates of return to evaluate the social benefits associated with the R&D investment. There are several empirical studies, especially focused on the agricultural sector, that provide robust findings of very high (20–60%) social rates of return (see Alston et al., 2009; Evenson, 2001). At the aggregate level, the social rate of return exceeds the private rate of return. Jaffe (1998) develops this type of model at the level of the US Advanced Technology Program (ATP) and shows that the social return could be increased by combining the positive effects of knowledge, markets, and network spillovers. Such approaches have the advantage that they quantify the economic benefits, and are useful for justifying existing public R&D programs. However, they do not add to our understanding of the process of generating economic benefits.

A number of evaluation studies analyze the various economic impacts of specific public R&D programs. The US ATP has conducted various evaluation studies using different methodologies (Ruegg and Feller, 2003). Apart from the above-mentioned social returns approach, ATP has used survey methods that exploit information on the activities, relationships, accomplishments, and licensing results of multiple actors. Also, case studies have been used to understand how and why certain developments, such as the dynamics of collaboration between different participants, emerged. Some case study investigations include quantification of the costs and benefits of the project. Econometric and statistical analyses provide information on the functional relationships between economic and social phenomena, and forecasts of economic effects, such as crowding out effects between public and private funds. Social network analyses have added to our understanding of how the structure of a project could increase the diffusion of new knowledge. Bibliometric studies have been applied to evaluate the number of publications and patent citations that have been generated by the program. Historical tracing helps to identify the linkages between a research project and its subsequent economic impacts.

Some of these methodologies have also been used to assess the European Framework and other programs (Georghiou and Roesner, 2000; Policy Research in Engineering Science and Technology PREST, 2002). Overall, this set of studies estimates a wide range of economic impacts, such as cost savings, rate of adoption of technologies, the impact of increased product quality on sales, efficiency of alliances, networking effects, human capital effects, impacts on firm's productivity, etc. Each study analyzes a particular type of economic impact based on specific hypotheses and analytical backgrounds on how the impact occurs. They take account of elements, such as research outputs, networking, cooperation, and learning processes.

Another stream of work focuses more generally on the influence of public research on industry R&D (Cohen et al., 2002). Using survey data, Mansfield (1991) shows that, over a 10-year period, 11% of new products and 9% of new processes would have been delayed at least one year in the absence of the public research conducted in the previous 15 years. Mansfield (1998) estimates that the value of the innovations that could not have developed without scientific results, accounts for 5% of total firm sales. Based on patent citations, Narin et al. (1997) show that between 1987 and 1994, the knowledge flow between US science and industry tripled. Well-known surveys, such as the Yale and the Carnegie Mellon Surveys, evaluate how public research impacts on industry innovation. They show that, with the exception of medicine, some chemical products, and electronics, universities and PROs have few direct effects on industry R&D (Klevorick et al., 1995). Cohen et al. (2002) show that firms use the following research outputs (in the order of their importance): research findings, new instruments and techniques, and prototypes. The authors highlight that public research results are transferred to industry via publications, informal interactions, conferences, and consulting. Patents and licenses as technology transfer mechanisms are useful in only a few industries.

In the context of outputs, Salter and Martin (2001) consider that publicly-funded research contributes to economic growth in several ways: by increasing the stock of knowledge, training skilled graduates, new scientific instrumentation and methodologies, development of networks, stimulation of social interactions, increased capacity for scientific and technological problem-solving, and creation of new firms. These contributions highlight the role played by research outputs and the channels used to exchange knowledge between public-funded research and industry. However, they focus mainly on the impacts of public research on product and process innovations, sales value, or industry R&D value. The main beneficiaries of public research in these analyses are industry or business actors.

2.2. Broader impact approaches

Bozeman (2003) argues that economic assessments of public research, whatever the method used, are based on the fundamental assumption that knowledge is a commodity valued at a price in a market. The economic transaction provides the basis for valuing the knowledge, albeit imperfectly. Economists do not estimate the value of knowledge based on the scope of its use. In the Public Value Mapping approach, knowledge gains value through its use and outcomes. The producers and users of the knowledge are crucial in the analysis. Moreover, "science outcomes are best understood in terms of the – 'knowledge value alliances' – that arise to generate, develop, and use scientific research. By this view, it is vital to understand research outcomes and the availability of scientific and technical human capital to produce research, but it is also important to understand other parties to the 'knowledge value alliance', including, as examples, government and private funding agents, end users, wholesalers, equipment and other scientific resource vendors, and so forth." (Bozeman, 2003: 13). Public Value Mapping considers outcomes, such as environmental quality and environmental sustainability, health care, and provision of basic needs, e.g., housing and food.

The idea of redefining impact assessment to include the broader societal returns is central to several papers in a Special Issue of *Research Evaluation* (2011). According to Donovan (2011), the definition of this broader impact should include the social, cultural, environmental, and economic returns: "how this broader impact is defined will determine how it is assessed" (Donovan, 2011: 176). To capture the broader societal benefits, "metrics-only approaches are behind the time, and state-of-the-art evaluations of research impact combine narrative with relevant qualitative and quantitative indicators" (*ibid*.: 176). The Payback Framework and the SIAMPI (Social Impact Assessment Methods for research and funding instruments through the study of Productive Interactions between science and society) are state-of-the-art evaluation approaches. They consider a variety of impacts, and take account of the diversity of the actors engaged in the knowledge generating process.

The Payback Framework was created to assess the outcomes of health research. It consists of two elements (Donovan and Hanney, 2011): a logic model of the research processes, and various categories of research paybacks (impacts). The logic model contributes to analyzing the 'story' of an innovation from topic identification, project specification, research process, and primary outputs of the research, to the various dissemination steps until the final outcomes. The dissemination and adoption phases tend to highlight the role played by intermediaries and beneficiaries. Various types of benefits are considered: academic benefits (publications, research reports, etc.), benefits for future research (development of research skills), benefits of policy and product development information, health sector benefits (improved health, improved equity in service delivery), and broader economic benefits.

The SIAMPI approach considers the 'productive interactions' between researchers and stakeholders as central to creating research with any kind of impact (Spaapen and Van Drooge, 2011). SIAMPI focuses on the interaction process in order to identify the relevance of the research, and how it is adopted, diffused, and applied, or not. Productive interactions are exchanges between researchers and stakeholders involved in achieving societal impacts (industry, public organizations, government, and the general public). The interaction is productive, because stakeholders make efforts to use and apply the research results. Societal impact stems from the process of knowledge creation, knowledge exchange, and knowledge circulation to achieve goals related to the development of sustainable societal development.

Finally, several agricultural PROs[2] have developed methodologies based on case studies to evaluate the various types of impacts generated by their research results. Their results underline that the research conducted affects a wide range of stakeholders in terms not only of economic impact but also environmental, health, and political impacts. Among these, the analysis by Douthwaite et al. (2003) (further taken up by Walker et al. (2008)) is interesting, because it introduces the notion of 'impact pathway'. An impact pathway captures the different stages of R&D from the basic research inputs to the final impacts, including the different research outputs and outcomes for different types of users. The impact pathway underlines the multiplicity of actors that contribute to the final impacts, and the various impact-generating mechanisms. The methodology also encompasses the diversity of impacts (economic productivity, social and distributional impacts, and environmental).

2.3. Our approach

Three main lessons can be established from the literature. First, it is now well established that assessment of research impact should deal with multiple dimensions covering not only economic, but also environment, health, social, and policy impacts. Second, innovation results from the activities and interactions of multiple actors. Understanding the contribution of a PRO to the impact requires paying attention to the intermediaries or beneficiaries (i.e., the innovation network) of the research results provided by the PRO. Third, case study methods are often used to analyze broader impacts and innovation networks. Each case study tends to be specific to the type of innovation analyzed and is a single story. As a consequence, it becomes difficult to get an overall picture of the impacts generated by one organization.

Our analysis aims at identifying the various impact patterns of a PRO, i.e., the various types of impact and the various ways by which the impact is generated. This work is applied to INRA, a French PRO involved in research and innovation related to agriculture, food, and the environment. Three sets of variables are used to define an impact pattern: research outputs, actors (beneficiaries or intermediaries), and the type of impacts (economic and non-economic). Despite its simplicity, this characterization of the impact pattern takes into account the multiple dimensions of impact and the multiple actors and outputs that contribute to the innovation. Note however that this notion of impact pattern does not enable to describe the process that generates impact from the research of the PRO and the multiple interactions within the innovation network. Our analysis is based on a set of 1048 innovation cases generated by INRA. It provides a good aggregated picture of the ability of INRA to generate societal impacts. By defining patterns, the objective is to identify the major regularities in the way research outputs are associated with beneficiaries or intermediaries and impacts.

[2] ACIAR – Australian Center for International Agricultural Research, (Pearce et al., 2006); EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA, 2013); USDA – U.S. Department of Agriculture, (Heisey et al., 2010); and BBSRC – Biotech and Biological Sciences Research Council in the UK (BBSRC, 2013).

Table 1
Modalities used to characterize each variable.

Variables		
Beneficiaries	Outputs	Impacts
B1: Public institutions (Spaapen and Van Drooge, 2011)	O1: Innovations embedded in technical objects (Cohen et al., 2002; Salter and Martin, 2001; Larédo, 1995; Colyvas et al., 2002; Kingsley et al., 1996)	I1: Economic competitiveness (Donovan and Hanney, 2011; Georghiou and Roessner, 2000 Salter and Martin, 2001; Bozeman and Melkers, 1993 Ruegg and Feller, 2003)
B2: Technical centres (Kingsley et al., 1996; Lyall et al., 2004)	O2: Innovations non embedded (Cohen et al., 2002; Salter and Martin, 2001; Larédo, 1995; Colyvas et al., 2002; Kingsley et al., 1996)	I2: Environment (Hermann et al., 2007; Donovan, 2011; Walker et al., 2008)
B3: Agricultural sectors[a], professional organizations (Lyall et al., 2004)	O3: Metrology, standards (Blind et al., 2010; Goluchowicz and Blind, 2011)	I3: Health (Donovan and Hanney, 2011; Bozeman, 2003)
B4: High technology industries (Spaapen and Van Drooge, 2011; Salter & Martin, 2001)	O4: Scientific advice (Grunwald, 2006)	I4: Social (Donovan, 2011; Bozeman, 2003; Ruegg and Feller, 2003; Molas-Gallart and Tang, 2011)
B5: Low technology industries (Lyall et al., 2004; Spaapen and Van Drooge, 2011; Salter and Martin, 2001)	O5: Coordination structures, institutions (Salter and Martin, 2001; Kamenetzky, 2013)	I5: Structuring a territory, a sector or market (Suh and MacPherson, 2007)
B6: Territories (Audretsch and Feldman, 2004; Pascucci and de-Magistris, 2011)	O6: Training (Salter and Martin, 2001; Martin and Salter, 1996; Kingsley et al., 1996; Bozeman and Kingsley, 1997; Kamenetzky, 2013)	I6: Public policy (Donovan and Hanney, 2011; Bell et al., 2011)
B7: Stakeholders, lobbies (Lyall et al., 2004; Kingsley et al., 1996)	O7: Banks, collections, databases (Martin and Salter, 1996)	I7: Maintaining of options for the future (Donovan and Hanney, 2011)
B8: Research and higher education (Spaapen and Van Drooge, 2011; Kingsley et al., 1996)		

[a] Agricultural sectors correspond to agricultural "filières".

3. Data

INRA is a mission oriented PRO specialized in the areas of agriculture, environment, and food. It has an annual budget of €880 million and employs 8500 individuals (including 1800 scientists). It accounts for most French public research in these areas. INRA's missions focus on the production of scientific knowledge, contributing to innovation, and scientific advice for policy makers. Its research is organized in 14 scientific departments that correspond to different disciplines including animal and plant sciences, environmental research, food technology and consumption, and social science[3].

The database we use is managed by INRA's Communication Department. It exploits several sources of information: significant research results since 2005, INRA press releases since 1996, and other communication media. Most of the information processed by the Communication Department is derived from a bottom–up selection process. This involves each INRA laboratory informing its scientific department of important results achieved each year, from which the most significant results are selected and passed to the Communication Department.

Significant results are defined based on a standardized description that includes the title, an abstract describing the innovation, the topic, the partners involved, and summary information related to its impact and intellectual property rights aspects. The original database contains 3589 entries, more than 2000 of which we excluded because of their academic topics. Significant research results linked to academic topics are related to more basic research which requires more time to produce market-related innovations. The final database used for this analysis contains 1048 entries. The distribution of the entries among scientific departments is uneven which might be explained by department size or the more or less applied nature of the research[4].

Three main qualitative variables were defined to characterize each of the entries: beneficiaries, outputs, and impacts. Seven or eight qualitative modalities are used to describe each variable, for a total of 22 modalities. These modalities are non-exclusive: one variable may be described by several modalities. Most modalities were chosen drawing on the evaluation studies literature, but we added some specific modalities related to agriculture. Table 1 presents the references used to build and select the modalities for each variable[5].

For beneficiaries, we mostly exploit the insights provided by Kingsley et al. (1996); Lyall et al. (2004); and Spaapen and Van Drooge (2011). Technical centers, high- and low-tech industries, and research and higher education organizations are classical innovation system actors. Stakeholders and lobbies are groups of actors with similar consumption habits, geographical specificities, or opinions (i.e., consumer groups, interest groups, NGOs, etc.). Public institutions are government organizations (i.e., ministries, administrations, funding agencies, etc.) that enforce the law, establish norms, ensure public interest, and provide public subsidies. We include the modality territories, as an agricultural specificity. Territories represent groups of actors with a common interest in locally valorized heritage, landscape, and "terroir" (e.g., Geographical Indication labeling).

The modalities for outputs are taken mainly from Salter and Martin (2001). Since we only consider entries that are linked to research results that are close to the market, we selected downstream outputs (i.e., tangible and intangible innovations, databases, etc.) and ignored upstream outputs (i.e., increased stock of useful knowledge). Coordination structures are organizational devices derived from INRA's research, such as joint R&D facilities (collaboration platforms, joint research units), consulting offices, or socio-professional networks. We include scientific advice, defined as accumulated multidisciplinary scientific knowledge pooled to provide inventories, diagnoses, and foresight to support public decision-making.

The impact modalities are inspired mainly by Donovan (2011); Donovan and Hanney (2011); and Larédo (1995). Since we are focusing on the agricultural sector, we include impacts on "territories", related to land-use, country and population planning aspects, or new ranges of products that influence existing markets (Geographical Indication labels have a structural impact on territories).

[3] There were no organizational changes at INRA in the period 1996–2010.

[4] The distribution of entries across departments ranges from 1% for Plant Biology (a small and fundamental research oriented department) to 15% for Science and Process Engineering of Agricultural Products (a large and applied research oriented department).

[5] There is no strict matching between our modalities and those in the literature: it is based on our interpretation.

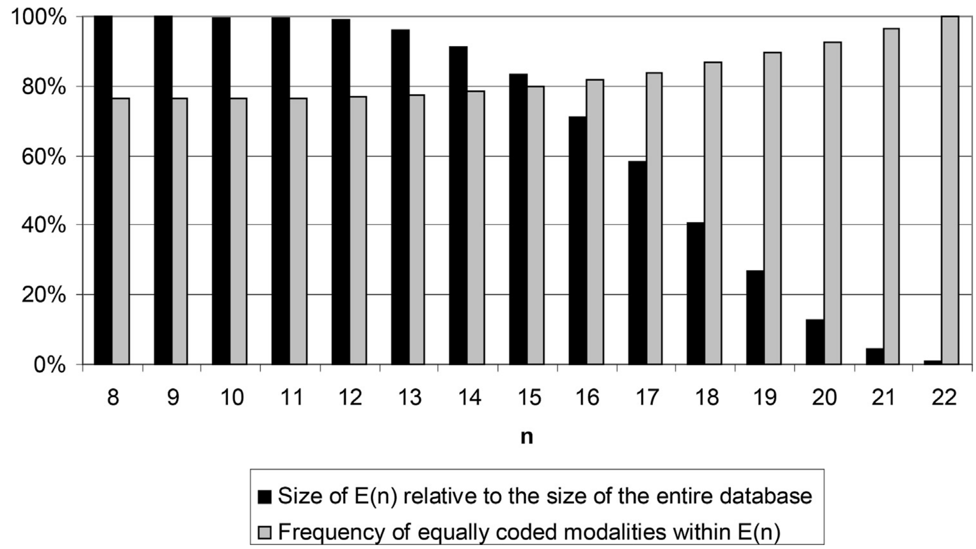

Fig. 1. Frequency of entries with equally coded modalities.

We also consider "maintaining options for the future" which corresponds to option values associated with the preservation of a good for its future use (e.g., biobanks which suggest opportunities for new research).

4. Methodology

The aim of the statistical analysis is to classify the 1048 entries into a limited number of classes representing our impact patterns. This requires monitoring the codification of the three variables, and applying an appropriate classification method to identify the main patterns.

4.1. Codification robustness

The modalities are non-exclusive and are dummies which score 0 for no modality and 1 otherwise. The 22 modalities are described in detail in precise guidelines written by the first codifier (one of the co-authors)[6]. The codification is based on careful reading of all the information available for each entry. In order to limit misinterpretation and bias, each entry was codified by three independent codifiers, all of whom are engineers and are knowledgeable about agronomy, agricultural sectors, policies, administrations, stakeholders, and the structure of INRA.

Codification robustness was analyzed by studying the frequency of the same codification by each of the three codifiers. We define $E(n)$ as the subset of entries with at least n modalities codified the same by the three codifiers. The sizes of $E(n)$ are denoted by the black bars in Fig. 1. The size of $E(n)$ decreases with n, because $E(n)$ is included in $E(n\text{-}1)$. All 1048 entries had at least eight modalities (out of 22) that were codified the same by all three codifiers. At the other extreme, for only 36 entries (4.5% of the database) were almost all the modalities (21 or 22) codified the same by all three codifiers. We also compiled the frequency of equally codified modalities within each subset $E(n)$ (grey bars in Fig. 1). When the entire database is considered ($E(8)$), 75% of the modalities received the same codification from all three codifiers.

To check the robustness of the codification further we tested different error rate levels. The error rate is defined as the probability that the codification assigned by the majority of codifiers is wrong[7]. For each modality, we test whether we can reject the hypothesis that the error rate is greater than a given value γ (see Appendix). Table 2 shows that when considering the entire database ($E(8)$), 14 modalities are codified with an error rate lower than 3%, four with an error rate of between 3% and 6%, and four with an error rate of over 6%. The error rate is generally higher for the modalities related to impacts, and is lower when considering subsets with higher numbers of equally codified entries. Two subsets are retained for the analysis: $E(14)$ and $E(17)$; in each subset, all the modalities are codified with an error rate lower than 6%, and 18 in subset $E(17)$ are codified with an error rate lower than 3%.

Table 2
Analysis of coding error rates.

Set or subset of entries	Entire database $E(8)$	$E(14)$	$E(17)$
Number of entries	1048 (100%)	953 (91%)	610 (58%)
Threshold to the error rate γ	Number of modalities coded with an error rate lower than the threshold		
<3%	14	14	18
<6%	18	22	22

4.2. The classification method (Partitioning Around Medoids)

We use the PAM-k Medoids (Partitioning Around Medoids) method which creates partitioned categories around central entries. The method is iterative and proceeds as follows. An initial given number of centers is chosen randomly among the entries, and the entries are then classified into the group corresponding to the closest center. New centers are defined based on the entries in each group[8]. Then the entries are reallocated to different groups, and so on. The iterations cease when the group structure is stable (i.e., average distance to the center of the group no longer decreases significantly). PAM methods differ depending on the rule adopted to identify the center of each category. The PAM-k Medoids is appropriate for our case because the center of the class corresponds to an

[6] The guidelines are available on request from the authors.

[7] This definition corresponds to a collective error rate and can be related to an individual error rate, as explained in Appendix.

[8] The center of a group is defined as the entry with the minimum average distance to all other of the group's entries.

Table 3
Distribution of modality frequencies within the $E(14)$ subset and the seven clusters.

	Classes title	*NE*1	*P*1	*NE*3	*P*2	*NE*2	Ba	SA	Subset E (14)
	Class size	162	196	135	131	106	64	159	953
Beneficiaries	Research	73	85	76	95	96	97	75	84
	Agric. sectors	79	91	71	31	16	44	65	62
	Public instit.	18	15	26	21	38	22	95	34
	Low tech.	23	20	17	35	9	2	5	17
	Technical Ct	19	25	19	4	5	8	19	16
	High tech.	12	19	13	33	7	30	4	16
	Stakeholders	4	6	8	9	12	8	16	9
	Territories	7	10	3	2	5	3	8	6
Outputs	Non-embedded Innov	91	5	83	6	85	13	22	43
	Technical product	9	80	8	77	6	3	3	31
	Scientific advice	9	9	16	1	8	5	62	17
	Coord. structures	7	9	7	24	7	11	20	12
	Biobanks	2	2	6	3	2	97	6	10
	Metrology	6	11	7	3	11	2	7	7
	Training	2	9	3	3	3	3	6	5
Impacts	Economic	97	92	81	82	20	73	72	77
	Environment	28	65	16	11	83	75	67	47
	Health	2	14	100	70	8	11	26	33
	Public policy	13	13	26	10	28	11	91	29
	Social	19	14	7	8	6	3	23	13
	Future options	7	2	7	9	13	94	0	12
	Struct. territory/market	12	15	6	9	8	2	17	11

existing entry whose interpretation is straightforward, and because it is less sensitive to outliers[9].

The PAM method is a recent development (Cardot et al., 2012; Park and Jun, 2009; Patel and Singh, 2013) and has been used mainly in computer science, statistics, bioinformatics, biotechnology, chemistry, and medicine. Its application in social science is limited. Maharaj and D'Urso (2010) identify patterns of yearly rates of change in the Gross Domestic Product of 22 developed countries. D'Urso and Massari (2013) identify human activity patterns and D'Urso et al. (2010) and Mackey and Poole (2008) analyze financial data and activities.

Our paper is the first to use the PAM method to describe heterogeneous innovation patterns. We consider it important for research managers and policy makers to have robust methods for analyzing the heterogeneity of research activities and actors or innovation project results. Funding and assessment agencies tend to build large datasets to produce indicators for research excellence (Research Excellence Framework in the UK) and to understand the results generated by public funding (Cordis in Europe, National Research Agency in France). Compiling statistical indicators based on these databases may not be sufficient to capture the richness of the data. Using robust clustering techniques, such as PAM to identify different patterns on the basis of all the available data (including qualitative data) could be a useful step towards compiling specific indicators for each pattern.

There are alternative classification methods. One of the most common in social science is hierarchical clustering. In an ascendant hierarchical clustering, each observation corresponds initially to a cluster, and pairs of clusters are merged as one moves up the hierarchy. The closest elements are merged in a cluster, according to a chosen distance (linkage criterion). The dendrogram represents the arrangement of the clusters produced by hierarchical clustering. The number of clusters in the partition is determined by cutting the dendrogram at a given level according to the selected precision. In our case, PAM is more appropriate because each cluster is easily interpreted based on its center representing a specific entry. It is also more flexible since entries can be reallocated among groups in any iteration, and it allows supervised classification (classifying new entries according to an existing stable classification, see Section 4.3).

4.3. Application of PAM clustering method to our data

Classification first requires the distance between entries to be defined. Here, the distance between two entries is defined as the number of modalities with different values. The PAM method has been applied to define classes of entries with similar profiles, based on a distance matrix.

To control for potential codification error, the PAM method was applied to a subset of entries $E(n)$ with low codification error. The choice of this subset is based on the trade-off between two criteria: on the one hand, we need to consider a subset large enough to obtain a representative sample and to avoid selection bias; on the other hand, we need to consider the robustness of the codification to avoid codification error leading to classification bias. The strategy considered to result in the best trade-off is to conduct the analysis in two steps, with a subset of different robustness degree for each step. First, we define the class centers on the basis of a narrow and robust subset, in order to limit bias due to mis-codification. We run the PAM method on the subset $E(17)$ which contains 610 entries with an error rate $\gamma < 3\%$ for 18 modalities. Second, to characterize each class, we use the centers defined previously to supervise the classification of a more representative set of entries, i.e., the subset $E(14)$ which has 953 entries (90% of our data set). The second classification is made using the closest neighbor algorithm: each entry is put into the class with the closest predefined centers, and centers are not re-calculated.

Several numbers of classes (between 5 and 11) were considered; we decided to use seven classes. This partitioning enables a fairly good reduction of the distances within classes, and good qualitative interpretation of classes in terms of impact patterns. Each of the seven classes denotes a pattern, but there can be different Variants of patterns within each class. To better describe the variability within each class, we built different sub-classes by performing PAM independently within each class.

[9] For instance, in the PAM-k means, the center is the calculated mean of the entries of the cluster, and does not necessarily correspond to an existing entry.

5. Results

5.1. Overview of the results

The modalities show very different frequencies (Table 3) in the set of 953 entries. The beneficiary Research is the most frequent modality, present in 84% of the entries, followed by Economic Impact (77%). At the other end are the outputs Training (5%) and Metrology (7%) with very low frequencies. In the following analysis of classes, we take care to differentiate modalities representative of a class from those over-represented in the whole set $E(14)$. Under-represented modalities in $E(14)$ may play an important qualitative role in some classes.

The final partitioning of the 953 entries we analyzed includes seven classes, each of which characterizes a specific impact pattern. Table 3 presents the classes, titled according to the types of output they include. NE refers to innovations Not Embedded in technical objects (methods, know–how) and accounts for 42% of the total sample of entries. *P* refers to Products (plant variety, device, software, or food product) and accounts for 34% of the entries. SA is Scientific Advice and accounts for 17% of the entries. Ba stands for bioBanks and accounts for 7% of the entries.

5.2. Descriptions of the seven classes

The class descriptions are based on the distribution of modalities within each class (Table 3) which allows us to name, interpret, and illustrate the seven classes. Each class corresponds to a specific impact pattern, but within each class, Variants of this pattern are identified according to the subclass analysis[10] (Figs. 2–8). The Variants introduce some variability within the class that might not emerge from consideration only of the class level. The Variants within classes always have some common variables, which guarantees a degree of homogeneity in the general interpretation of the class.

The seven classes differ, in particular, according to the combinations of impacts. The impacts match the diversity of INRA's missions. Two classes (see Section 5.2.1) represent innovations related to INRA's historical mission of fostering agricultural sector competitiveness and improving the environment. Two other classes (see Section 5.2.2) relate to innovations addressing a more recent mission related to public health issues and two patterns (see Section 5.2.3) correspond to INRA's efforts to preserve the environment and manage natural resources. Finally, INRA aims to provide scientific advice (see Section 5.2.4) to support policy decision-making. The remaining subsections describe the classes according to types of outputs and related beneficiaries representative of individual patterns.

5.2.1. Innovations supporting the agricultural sector and generating economic and environmental impacts (NE1, P1)

The distinction between the two classes (NE1, P1) is related mainly to the types of outputs (products vs. methodologies) developed by INRA's researchers. These outputs promote increased economic competitiveness based on intensive use of improved inputs (e.g., pesticide and vaccine), and more productive livestock and crops. This corresponds to the first pillar of the European Common Agricultural Policy (CAP) of supporting of farmers' productivity.

[10] Variants are identified for each class (17 in total), and the proportion of entries in that class that they cluster is presented in the last column of Figs. 2–8. For some classes, a few fringe entries did not comply with any of the variants and are excluded.

5.2.1.1. Class NE1: methodological breakthrough supporting the economic competitiveness of the agricultural sector (17% of entries). NE1 (Fig. 2) clusters innovations not embedded in technical objects (91% of the entries), that generate an economic impact (97%). These innovations can be methods, processes, protocols, decision-supporting tools, or mathematical models. They target agricultural sectors in the broad sense (farmers, seed producers, technical centers, advisory offices, and agro-food industry). Their economic impact includes more competitive farming systems, increased yields, and creation of added value along the agricultural sector value chain.

A typical example of these innovations and their impact is the implementation of new techniques for artificial insemination which require new breeding practices (methods) to be designed. Artificial insemination generates broad changes at the farm level, and favors the diffusion of genetic progress, enabling increased production performance, and thus, producing economic impacts.

The three Variants differ mainly in the vector of impact generated by the agricultural sector. The economic impact dominates in this class (Variant 1 clusters entries that produce only economic impact), but in some cases, might be associated with environmental (Variant 2) or social impacts (Variant 3).

5.2.1.2. Class P1: embedded technologies that have an impact on agricultural sector economic competitiveness and the environment (21% of entries). Innovations in class P1 (Fig. 3) are mostly new products (80% of entries) targeting agricultural sectors. They include equipment, food products, and pest treatment solutions. Their economic impact dominates (92%), but they also target environmental, social, and territorial issues.

A representative case is the creation of a new seed variety (product) with increased natural resistance to various fungi which contributes to increasing yields (economic impact) and reducing use of alternative chemical treatments (environmental impact).

Variant 2 is the dominant sub-class linking technical products to low-tech firms and the agricultural sectors, generating economic as well as environmental impacts. Variant 1 corresponds to new products, such as new plant varieties and animal breeds which benefit farmers and technical centers, generating economic, social, and territorial impacts. Variant 3 differs from the other two Variants mainly in terms of output which is scientific advice and metrology (standards, Geographical Indication labels).

The two patterns identified here are found in other sectors. Products and methodologies generated by public research often contribute to the economic growth of firms in various sectors (Rosenberg, 1992; Salter and Martin, 2001; Von Hippel, 1987).

5.2.2. Innovations devoted to the agricultural sectors generating health and economic impacts (NE3, P2)

The distinctions between the two classes in this group (NE3, P2) are related mainly to types of outputs (methodologies vs. products) and beneficiaries (agricultural sectors vs. industry and research). Both generate health and economic impacts. Major sanitary crises, such as so called mad cow disease or BSE, have contributed to more research on health issues.

5.2.2.1. Class NE3: methodological development for agricultural sectors and public institutions impacting health and the economy (14% of entries). Class NE3 (Fig. 4) groups methodologies (83% of entries) generating health (100%) and economic (81%) impacts. Health impacts comprise human and animal health, food quality and safety, and aspects of well-being. Agricultural sectors and technical centers play key roles in the diffusion of these methodologies.

A representative example is spectroscopy analysis of adipose tissues, a method that provides information on feed diets for lambs, which has sanitary and economic implications.

Variants of impact pattern			**Examples of entry title**	**% of entries of the class**
Outputs	Beneficiaries	Impacts		
	Agri. sectors	Economic	Improved practices for large scale diffusion of Artificial Insemination (Variant 1)	43
Innovation not embedded in technical products	Agri. sectors and Technical Centers	Economic and Environment	Irrigation management (Variant 2)	30
	L. Tech	Economic and Social	A modeling tool to manage seasonal work in response to food sector demand (Variant 3)	12

Fig. 2. Impact pattern NE1.

Variants of impact pattern			Examples of entry title	% of entries of the class
Outputs	Beneficiaries	Impacts		
Technical products	Agri sectors and Technical Ct	Economic and social and struct. territory/market	Carmine, a new lettuce for year 2000 (Variant 1)	8
	Agric sectors and/or L. Tech		Plants resistance to parasites : new arms for biological control (Variant 2)	69
cient. advice and metrology	Agric sectors	Economic and/or Environment	Climate change: new drought-resistant tree species (Variant 3)	11

Fig. 3. Impact pattern P1.

Variants of impact pattern			Examples of entry title	% of entries of the class
Outputs	Beneficiaries	Impacts		
Innovation non embedded in technical products	Agri. sectors and/or Technical Ct	Economic and Health	Reflectance spectroscopy for the authentication of lambs' diet (Variant 1)	64
	Public instit. and Agri. sectors	Economic and health and public policy	Mathematics and computer tools for modeling and prediction of dynamic processes in animal epidemiology (Variant 2)	19

Fig. 4. Impact pattern NE3.

Variant 1 is the dominant pattern: a methodology applied in the agricultural sector allows health and economic benefits. The difference in Variant 2 is that the methodologies also benefit public institutions (i.e., government organizations) and influence public policy. For example, epidemiological research informed the design of disease control policies.

5.2.2.2. Class P2: embedded technologies and coordination structures for private firms, and research generating health and economic impacts (14% of entries). P2 (Fig. 5) clusters products (77% of entries) and coordination structures (24% of entries), designed for public research institutions (95%) and for low- (35%) and high- (33%) tech firms. They generate both economic and health impacts.

A typical example is natural food coloring extracted from apple juice. Easy to use in food processes (economic impact on low-tech firms) because its structure is stable, it contains healthy antioxidant properties (health impact).

Variants of impact pattern			Examples of entry title	% of entries of the class
Outputs	Beneficiaries	Impacts		
Technical products	L. or H. Tech and Research	Economic and Health	Getting a yellow hydrosoluble food color from an apple (Variant 1)	76
Coordination structures		Economic	Technological platform for the structural analysis of molecules (Variant 2)	17

Fig. 5. Impact pattern P2.

Innovations of Variant 1 mainly affect health and the economy through the provision of products used by the private sector and research institutions. Variant 2 differs from Variant 1 mainly in terms of outputs. Coordination structures (joint research units, collaboration platforms, etc.) established by INRA help to foster competitiveness through collaborative research, and networking between public research institutes and firms. In other words, this P2 impact pattern relies on two types of devices (products and organizational devices) that serve two actors central in the innovation system (firms and research institutions) and also their networking opportunities to generate economic and health impacts.

Compared to the two "agriculture-competitiveness" patterns related to INRA's historical mission (cf. 5.2.1), the two "health-economic" patterns described here are related to more recent efforts to address public health issues. The outputs and beneficiaries are rather similar among the four patterns. However, P2 underlines the important role of INRA in encouraging networking between public research actors and firms (Callon, 1994).

5.2.3. *Innovations impacting the conservation of natural resources (NE2, Ba)*

The difference between the two classes below (NE2, Ba) lies in the types of outputs (methodologies vs. biobanks). These patterns correspond to recent INRA missions formulated in response to growing global interest in reducing environmental impacts. The interest of public research in natural resource management (NRM) emerged in the 1990s (Renkow and Byerlee, 2010). The second pillar of the CAP which focuses on environmental conservation and sustainable and territorial development, has contributed to the emergence of these types of innovations.

5.2.3.1. Class NE2: methodological breakthroughs that benefit PROs are related to current and future environmental issues (11% of entries). Class NE2 (Fig. 6) is characterized by methodological breakthroughs (85%) for public research actors (96%) impacting environment. Environmental impacts (83%) are understood broadly as encompassing biodiversity, pollution, waste and energy management, and climate change.

For example, the successful adaptation of existing IVF (in vitro Fertilization) methods to deer opens up possibilities for supporting endangered species. The implanting of embryos in surrogates could help to sustain rare species thereby maintaining biodiversity (environmental and future option impact). PROs are central actors in these efforts.

Variant 1 is dominant and shows the central role of PROs in generating, developing, and diffusing these methods which generate environmental impacts and contribute to supporting options for the future through conservation and prediction solutions, and biotechnologies. Compared to Variant 1, Variant 2 includes public institutions as beneficiaries. The methodologies improve environmental performance and support public policy actions to PROs' research activities (Variant 2).

5.2.3.2. Class Ba: management of biobanks for public and private R&D to ensure options for the future and preserve the environment (7% of entries). Class Ba (Fig. 7) clusters innovations based on the management of biobanks, and plant and animal collections which conserve existing resources (97%). The exploitation of this output by public research actors has a strong influence on future opportunities (94%) to preserve resources and improve fundamental knowledge. The use of biobanks by the high-tech and agro-industries generates environmental, and to a lesser extent economic impacts.

The collection of natural and cultivated sunflower varieties is a representative case. It was established in the 1970*s* through the joint efforts of INRA and various seed breeders. It has enabled the creation of genetic source populations for breeding activities (high-tech firms) to suit future demand.

Variants of impact pattern			Example of entry title	% of entries of the class
Outputs	Beneficiarie s	Impacts		
Innovation non embedded in technical products	Research	Environment and/or Economic or Future options	Birth of a *sika* deer calf after transfer of IFV embryo (Variant 1)	75
	Research and Public instit.	Environment and Public policy	Algorithm for the calculation of nitrous oxide soil emissions (Variant 2)	25

Fig. 6. Impact pattern NE2.

Variants of impact pattern			Examples of entry title	% of entries of the class
Outputs	Beneficiaries	Impacts		
Biobanks	Research and Agric Sectors or H. Tech	Economic and Environment and Future options	Collection of sunflowers (Variant 1)	75
	Research	Future options	Information system to manage and analyze Arabidopsis thaliana transcriptome : CATdb (Variant 2)	25

Fig. 7. Impact pattern Ba.

Variant 1 and Variant 2 differ in beneficiaries and impacts. The use of biobanks by the public research sector is aimed at maintaining future research options (Variant 2). High-tech industries or agricultural sectors use the material for R&D projects to achieve economic and environmental objectives (Variant 1).

These two "preservation of natural resources" patterns are much more specific than the "agriculture-competitiveness" and "health-economic" patterns, and involve a central role of public research actors in building and using methodologies and biobanks to explore new research paths and address environmental concerns. INRA's use of biobanks relates mainly to plant genetics and biological resources to preserve biodiversity. They act as insurance against environmental crises, and preserve future innovative product and method development opportunities. They are used also by industry to obtain economic benefits while taking account of environmental concerns.

5.2.4. Scientific advice to inform public decisions (SA)

Scientific advice to inform policy decision making is a specific responsibility of PROs and involves unbiased advice in situations of scientific and technical controversy. This advice draws on the organizations' multidisciplinary skills which are rarely embodied in single private actors.

5.2.4.1. Class SA: scientific advice and coordination structures to inform public decision-making about sustainable development issues (17% of entries). Class SA (Fig. 8) consists of scientific advice (62% of entries) and to a lesser extent of coordination structures (20%) which impact on public policy (91%). Scientific advice is exploited by public institutions (ministries, national parks, funding agencies) and the agricultural sector (including technical centers).

For instance, INRA advised the French agriculture and environment ministries on agriculture and biodiversity (public institutions). It synthesized knowledge on the impact of agriculture on natural biodiversity (environmental impact) and the services that biodiversity may provide to agriculture. It allows ways to account for biodiversity in policy formulation (political impact).

The Variants differ in outputs and combinations of impacts. Scientific advice can be seen as an output based on a body of long-term accumulated multidisciplinary knowledge, pooled to produce inventories, diagnoses, and foresight. Coordination structures (observatories, joint research units) are organizational facilities that enable the pooling and integration of scattered knowledge. These two types of outputs are temporary configurations able to provide scientific background to political decisions on environmental and social issues.

Scientific advice is designed to inform the decision-making of public institutions and agricultural sectors. In addition to its impact on public policy, scientific advice has environmental impacts when it is related to natural resources management, renewable, or climate change for instance (Variant 1). Its social impacts (Variant 2) can be seen in relation to agricultural systems, rural organizations, and job market issues. Variant 3 is related to coordination structures, and benefits professional stakeholders, technical centers, and public institutions, and thus impacts on public policy.

Variants of impact pattern			Examples of entry title	% of entries of the class
Outputs	Beneficiaries	Impacts		
Scientific advice		Environment and Public policy	Scientific advice « agriculture and biodiversity » (Variant 1)	48
	Public instit. and/or Agri sectors and/or Technical Ct	Social and Public Policy	Public policies evaluation in employment and rural development (Variant 2)	12
Coord. Structure		Public policy	Development of AGENAE scientific consortium (Variant 3)	40

Fig. 8. Impact pattern Ex.

INRA contributed to the CAP measures. For instance, its scientific advice influenced the introduction of agro-environmental and other specific measures related to agro-forestry, which anticipated the consequences of the 2013 CAP reform.

6. Discussion and Conclusion

Based on a PAM classification of more than a thousand innovation entries we extracted seven impact patterns characterizing INRA's ability to fulfill its various missions. Each impact pattern has a specific configuration of outputs and beneficiaries which generate particular impact vectors. Our results can be summarized as follows:

– INRA's historical mission of supporting the agricultural sector competitiveness is sustained by two "agriculture-competitiveness" patterns. One is characterized by methods and the other by products, and both are useful to various actors in the agricultural value chain to generate economic and environmental impacts;
– INRA's more recent missions related to health as a result of major sanitary crises, exhibit two "health-economic" patterns. The first refers to methodologies diffused to the agricultural sectors. The second involves products and coordinating structures used by the industry and public research actors to create favorable synergies;
– INRA's environmental conservation mission involves two "preservation of natural resources" patterns. One is described by methodologies developed by and for PROs to respond to environmental concerns. The other focuses on building biobanks for exploitation by public researchers to preserve options for the future, or by industry to meet economic and environmental objectives;
– INRA also seeks to inform public decision-making by providing scientific advice. This pattern is an original result rarely highlighted in the literature on the impact of public research.

In the remaining parts of the paper we relate our results to the literature in order to highlight our contribution and offer some research perspectives.

6.1. *Analyzing impacts at the aggregate level of a PRO*

Compared to the literature reviewed in Section 2, our results exhibit the following originality and strength. Our analysis considers multiple types of impacts including economic, environmental, health, and policy related. We have shown that in most cases, INRA has an impact on society through more than just the economic dimension, although this is the most frequent. More importantly, our results underline that each impact pattern includes a combination of at least two major impacts. Not all types of impact combination are important. The most relevant combinations are economics and environment, health and economics, public policy and environment, and environment and preserving options for the future.

Compared to broader impact approaches, our analysis provides patterns at the aggregate level of an organization, based on a large sample of research events and innovations generated by INRA's research groups. To our knowledge, there are very few analyses based on such a large sample of events. The interest of our classification is to provide an overall and representative picture of a PRO's innovation patterns which could be complemented by in-depth case studies to provide more precise analysis of the mechanisms linking outputs, actors, and impacts.

6.2. *Implications of our results, and research perspectives*

This paper should be considered a first step in a wider evaluation exercise aimed at conducting case studies to quantify the various impacts, and to understand the mechanisms generating them. Our results underline the variety of impact combinations generated at the level of the organization, but do not provide information on their intensity. The typology does not analyze skewness, e.g., are a few innovations responsible for the largest impacts? This is an important limitation compared to studies evaluating the economic impact of public research. Using quantitative methods, these approaches confirm a skewed distribution of economic benefits where only a few licensed patents generate high royalty revenue (Mowery and Sampat, 2005) and a few publicly-funded projects generate high turnover (Georghiou, 1999).

One of the first steps in many evaluation studies, is building a representative sample of projects to evaluate. This sampling is complicated in a PRO organized around departments and activities rather than projects. To evaluate the societal impact of a PRO based on case studies, it is important to select cases which cover the diversity of the activities conducted, and to select cases relevant to analyzing impacts at the general level of the institution. The classification proposed in this paper addresses these selection criteria. In our evaluation study of INRA, we selected the activities to be evaluated by discussing the seven impact patterns identified with the scientists in charge of INRA's 14 scientific departments. The completion of more than 30 case studies will allow us to complement this important work with qualitative (i.e., information gathered on the mechanisms linking output, actors, and impacts) and quantitative (impact measurement) analyses of major innovations. One of the important challenges related to the case studies will be quantifying the effects.

Acknowledgments

The authors would like to thank Bénédicte Gouraud in INRA's Evaluation Office for her assistance with the statistical analysis. We also thank Laurence Colinet, Pierre-Benoît Joly, Philippe Larédo, Irwin Feller, Monica Gaughan, and Barry Bozeman for advice and insights. This paper benefited from comments from the participants at two conferences (Evaluation of STI policies, instruments and organization Conference, Vienna, 2013), and R&D Management Conference, Grenoble, 2012). The analysis would have been impossible without financial support from INRA (France) through the ASIRPA project (Socio-economic Analysis of the diversity of Impacts of Public Agricultural Research).

Appendix.

A.1. Testing the individual error rate for a given modality

We define α as the probability that one codifier makes a mistake in codifying one of the modalities. We can build the statistic Y as the number of entries where the codification modality is the same. Y follows a binomial distribution $B(N, \beta)$, N is the number of entries, and β is the probability that the three codifiers agree. β is related to α as follows:

$$\beta = \alpha^3 + (1 - \alpha)^3$$

The hypotheses of the test are:

– H0: the error rate of a codifier is greater than or equal to α.
– H1: the error rate is less than α.

Table 4
Detailed results of codification error rate for each of the modalities.

Error rate (x)		Initial set 1048 forms $E(8)$		Quality set 953 forms $E(14)$		Robust set 610 forms $E(17)$	
Individual α	Collective γ	Modalities[a]	Number of modalities with an error rate $>x$	Modalities[a]	Number of modalities with an error rate $>x$	Modalities[a]	Number of modalities with an error rate $>x$
≥15%	≥6%	I1 I3 B8 I6	4		0		0
10%≤<15%	3%≤<6%	O2 B3 I4 B1	4	I1 I3 B8 B3 O2 I6 I4 B1	8	I1 I3 B8 B3	4
<10%	<3%	I5 I2 O1 B4 B5 B7 I7 O4 B6 O3 O5 O6 B2 O7	14	I5 I2 O1 B4 B5 B7 I7 O4 B6 O3 O5 O6 B2 O7	14	O2 I6 I4 B1 I5 I2 O1 B4 B5 B7 I7 O4 B6 O3 O5 O6 B2 O7	14

[a] With decreasing value of the error rate.

The null hypothesis H0 is rejected if the value of Y is large. More precisely, we reject H0 if the normalized statistic is larger than the chosen quantile of the normal distribution.

$$U = \frac{Y - N\beta}{\sqrt{N\beta(1-\beta)}}$$

A.2. Collective error rate for a given modality (γ)

We define γ as the probability that the codification defined by the majority of the codifiers is wrong. γ can be compiled as a function of α. If we define C as the set of the three codes for the modality, M as the true value of the modality, and p the probability that the true value of M is 1, we have:

$$\begin{aligned}\gamma\ =\ & \Pr\{C=(0,0,0)\text{and}M=1\} + \Pr\{C=(1,1,1)\text{and}M=0\}\\ & +\Pr\{C=(1,0,0)\text{and}M=1\} + \Pr\{C=(1,1,0)\text{et}M=0\}\\ & =\alpha^3 p+\alpha^3(1-p) + 3\alpha^2(1-\alpha)p+3\alpha^2(1-\alpha)(1-p)\\ & =\alpha^2(3-2\alpha)\end{aligned}$$

A.3. Detailed results for codification error rate for each of the modalities

Table 4

References

Alston, J.M., Pardey, P.G., James, J.S., Andersen, M.A., 2009. The economics of agricultural R&D. Ann. Rev. Resour. Econ. 1, 537–565, http://dx.doi.org/10.1146/annurev.resource.050708.144137.

Audretsch, D.B., Feldman, M.P., 2004. Chapter 61 knowledge spillovers and the geography of innovation. In: Henderson, J.V., Jacques-François, T. (Eds.), Handbook of Regional and Urban Economics. Elsevier, pp. 2713–2739 http://www.sciencedirect.com/science/article/pii/S157400800480018X

Bell, S., Shaw, B., Boaz, A., 2011. Real-world approaches to assessing the impact of environmental research on policy. Res. Eval. 20 (3), 227–237.

Blind, K., Gauch, S., Hawkins, R., 2010. How stakeholders view the impacts of international ICT standards. Telecommun. Policy 34 (3), 162–174.

BBSRC, Our impact [www document], BBSRC (2013) http://www.bbsrc.ac.uk/research/impact/impact-index.aspx (accessed 11.05.13).

Bozeman, B., et al., 2003. Public value mapping of science outcomes: theory and method. In: Sarewitz, D. (Ed.), Knowledge Flows & Knowledge Collectives: Understanding the Role of Science & Technology Policies in Development, 2, pp. 3–48.

Bozeman, B., Kingsley, G., 1997. R&D value mapping: a new approach to case study-based evaluation. J. Technol. Transfer 22 (2), 33–41.

Bozeman, B., Sarewitz, D., 2011. Public value mapping and science policy evaluation. Minerva 49, 1–23.

Callon, M., 1994. Is science a public good? Sci. Technol. Hum. Values 19, 345–424.

Cardot, H., Cénac, P., Monnez, J.-M., 2012. A fast and recursive algorithm for clustering large datasets with k-medians. Comput. Stat. Data Anal. 56, 1434–1449, http://dx.doi.org/10.1016/j.csda.2011.11.019.

Cohen, W.M., Nelson, R.R., Walsh, J., 2002. Links and impacts: the influence of public research on industrial R&D. Manage. Sci. 48, 1–23.

Colyvas, J., et al., 2002. How do university inventions get into practice? Manage. Sci. 48 (1), 61–72.

Donovan, C., 2011. State of the art of assessing research impact: introduction to a special issue, research evaluation. Res. Eval. 20, 175–179.

Donovan, C., Hanney, S., 2011. The payback framework explained. Res. Eval. 20, 181–183.

Douthwaite, B., Kuby, T., van de Fliert, E., Schulz, S., 2003. Impact pathway evaluation: an approach for achieving and attributing impact in complex systems. Agric. Syst. 78, 243–265, http://dx.doi.org/10.1016/S0308-521X(03)128-8.

D'Urso, P., Cappelli, C., Di Lallo, D., Massari, R., 2010. Clustering of financial time series. Phys. A: Stat. Mech. Appl. 392 (9), 2114–2129, http://dx.doi.org/10.1016/j.physa.2013.01.027.

D'Urso, P., Massari, R., 2013. Fuzzy clustering of human activity patterns. Fuzzy Sets and Systems 215, 29–54, http://dx.doi.org/10.1016/j.fss.2012.05.009.

EMBRAPA's Social Report 2012, Distrito Federal (DF): Embrapa, Communication Secretariat, Strategic Management Secretariat, Brasilia (2013).

Etzkowitz, H., 1998. The norms of entrepreneurial science: cognitive affects of the new university-industry linkages. Res. Policy 27, 823–834.

E. Evenson, Economic impacts of agricultural research and extension. In: B.L.Gardner, G.C. Rausser (Eds.), Handbook of Agricultural Economics, 2001; 573-628.
Georghiou, L., 1999. Socio-economic effects of collaborative R&D—European experiences. J. Technol. Transfer 24, 69–79.
Georghiou, L., Roesner, D., 2000. Evaluating technology programs: tools and methods. Res. Policy 29, 657–678.
Goluchowicz, K., Blind, K., 2011. Identification of future fields of standardisation: an explorative application of the Delphi methodology. Technol. Forecasting Social Change 78 (9), 1526–1541.
Grunwald, A., 2006. Scientific independence as a constitutive part of parliamentary technology assessment. Sci. Public Policy 33 (2), 103–113.
Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., Trow, M., 1994. The New Production of Knowledge: Dynamics of Science and Research in Contemporary Societies. SAGE Publications Ltd., London.
Heisey, P.W., King, J.L., Rubenstein, K., Bucks, D.A., Welsh, R., 2010. Assessing the Benefits of Public Research Within an Economic Framework The Case of USDA's Agricultural Research Service (No. Economic Research Report Number 95). USDA Economic Research Service.
Hermann, B., Kroeze, C., Jawjit, W., 2007. Assessing environmental performance by combining life cycle assessment, multi-criteria analysis and environmental performance indicators. J. Cleaner Prod. 15, 1796–1987.
Jaffe, A., 1998. The importance of spillovers in the policy mission of the advanced technology program. J. Technol. Transfer 23, 11–19.
Kamenetzky, J.R., 2013. Opportunities for impact: statistical analysis of the national science foundation's broader impacts criterion. Sci. Public 40, 72–84, http://dx.doi.org/10.1093/scipol/scs059.
Kingsley, G., Bozeman, B., Coker, K., 1996. Technology transfer and absorption: an R&D value-mapping approach to evaluation. Res. Policy 25, 967–995.
Klevorick, A., Levin, R., Nelson, R.R., Winter, S., 1995. On the sources and significance of inter-industry differences in technological opportunities. Res. Policy 24, 195–205.
Larédo, P., 1995. Structural effects of EC RT & D programmes. Scientometrics 34, 473–487.
Lyall, C., Bruce, A., Firn, J., Firn, M., Tait, J., 2004. Assessing end-use relevance of public sector research organisations. Res. Policy 33, 73–87.
Mackey, S., Poole, B., Spring 2008. Estimating hedge fund investment sub-styles. J. Bus. Econ. Stud. 14 (1), 1–18.
Maharaj, E.A., D'Urso, P., 2010. A coherence-based approach for the pattern recognition of time series. Phys. A: Stat. Mech. Appl. 389, 3516–3537, http://dx.doi.org/10.1016/j.physa.2010.03.051.
Mansfield, E., 1991. Academic research and industrial innovation. Res. Policy 20, 1–12.
Mansfield, E., 1998. Academic research and industrial innovation: an update of empirical findings. Res. Policy 26, 773–776.
Martin, B., 2011. The research excellence framework and the impact agenda: are we creating a frankenstein monster? Res. Eval. 20, 247–254.
Martin, B.R., Salter, A., 1996. The Relationship Between Publicly Funded Basic Research and Economic Performance. University of Sussex: Science Policy Research Unit.
Molas-Gallart, J., Tang, P., 2011. Tracing 'productive interactions' to identify social impacts: an example from the social sciences. Res. Eval. 20 (3), 219–226.
Mowery, D.C., Sampat, B.N., 2005. Universities in national innovation systems. In: In Oxford Handb. Innovation. Oxford University Press, New-York, pp. 209–239.
Narin, F., Hamilton, K.S., Olivastro, D., 1997. The increasing link between U.S. technology and public science. Res. Policy 26, 317–330.
Park, H.-S., Jun, C.-H., 2009. A simple and fast algorithm for K-medoids. Expert Syst. Appl. 36, 3336–3341, http://dx.doi.org/10.1016/j.eswa.2008.01.039.
Pascucci, S., de-Magistris, T., 2011. The effects of changing regional agricultural knowledge and innovation system on Italian farmers' strategies. Agric. Syst. 104 (9), 746–754.
Patel, A., Singh, P., 2013. New approach for K-mean and K-medoids algorithm. Int. J. Comput. Appl. Technol. Res. 2, 1–5.
Pearce, D., Monck, M., Chadwick, K., Corbishley, J., 2006. Benefits to Australia from ACIAR-funded research (No. 39), Impact Assessment Series Report. Australian Centre for International Agricultural Research, Canberra.
Policy Research in Engineering Science and Technology, 2002. Assessing the Socio-economic Impacts of the Framework Programme. University of Manchester.
Renkow, M., Byerlee, D., 2010. The impacts of CGIA research: a review of recent evidence. Food 35, 391–402, http://dx.doi.org/10.1016/j.foodpol.2010.04.006.
Rosenberg, N., 1992. Scientific instrumentation and university research. Res. Policy 21, 381–390.
Ruegg, R., Feller, I., 2003. A Toolkit for Evaluating Public R&D Investment: Models, Methods, and Findings from ATP's First Decade. National Institute of Standards and Technology, Gaithersburg (Grant/Contract Report).
Salter, A., Martin, B.R., 2001. The economic benefits of publicly funded basic research: a critical review. Res. Policy 30, 509–532, http://dx.doi.org/10.1016/S0048-7333(00)91-3.
Spaapen, J.M., Van Drooge, L., 2011. Introducing productive interactions in social assessment. Res. Eval. 20, 211–218.
Suh, J., MacPherson, A., 2007. The impact of geographical indication on the revitalisation of a regional economy: a case study of 'Boseong' green tea. Area 39 (4), 518–527.
Von Hippel, E., 1987. Co-operation between rivals: informal know-how trading. Res. Policy 16, 291–302.
Walker, T., Maredia, M., Kelley, T., Rovere, R.L., Templeton, D., Thiele, G., Douthwaite, B., 2008. Strategic Guidance for Ex Post Impact Assessment of Agricultural Research. Report Prepared for the Standing Panel on Impact Assessment. CGIAR Science Council, Science Council Secretariat, Rome Italy.